12. Using the 1990–1993 Full Panel Longitudinal Research Files

This chapter discusses procedures for working with data from the full panel longitudinal research files for the 1990 through 1993 Panels of the Survey of Income and Program Participation (SIPP). Because the full panel longitudinal research file for the 1996 Panel was still under development at the time this chapter was written, it is not yet possible to describe procedures for using that file. A revised version of this chapter will be available once the longitudinal research file for the 1996 Panel is released to the public.

The chapter begins by describing the documentation that accompanies the full panel public use files obtained from the Census Bureau. The discussion then turns to the data files themselves. The data file structure is described, and detailed explanations are provided about how to use the longitudinal research files when performing common tasks, including:

- Realigning the data by calendar month;
- Using the monthly interview status variables;
- Identifying persons, households, families, and program units;
- Working with the unearned income data;
- Understanding the effects of topcoding;
- Using imputation flags; and
- Identifying states and metropolitan areas.

Before reading this chapter, users should read Chapter 9 for an introduction to Section II. Analysts using only one longitudinal research file should also read about the use of sample weights (Chapter 8) and the computation of standard errors (Chapter 7). Those planning on merging data from a longitudinal research file to data from the core wave or topical module files should read Chapter 10 for information about the core wave files, Chapter 11 for information about the topical module files, and Chapter 13 for information about linking SIPP public use files.

This chapter focuses on the longitudinal research files. It is written so that it can be used independently of the chapters describing the core wave files and topical module files. Although there are many similarities across the three types of files, important differences do exist. Because those differences are sometimes subtle, users familiar with the core wave and topical module files should read this chapter carefully, paying close attention to information about variable

names and file structures. Table 9-2 summarizes the differences between the core wave, topical module, and longitudinal research files.[1]

Using the Technical Documentation of the 1990–1993 Longitudinal Research Files

Each data file received from the Census Bureau comes with a set of technical documentation and a data dictionary. The technical documentation includes:

- The paper survey instrument;
- A glossary of selected terms;
- A cross-walk, mapping reference months into calendar months for each rotation group;
- A source and accuracy statement describing the sample weights and the computation of standard errors; and
- User Notes.

The survey instrument is vital to understanding what questions were asked, how they were asked, the order in which they were asked, to whom they were asked, and the way in which the answers were recorded. Some questions employ skip patterns (Chapter 3), so users should pay particular attention to which questions were skipped for which respondents. These skip patterns are best understood by consulting the survey instruments.[2]

The source and accuracy statements provide information about the weights on the files, when and how to make adjustments to the weights, and one approach to computing standard errors for some common types of estimates. More detailed discussions of those topics are provided in Chapters 7 and 8 of this Guide.

The data dictionary provides a detailed description of each variable on the file. It describes four aspects of each variable:

1. The definition;
2. The sample universe of the corresponding survey question;

[1] Some of this information will change once the 1996 longitudinal research file becomes available. At that time, this guide will be updated to reflect the differences.

[2] With the introduction of CAI (computer-assisted interviewing) in the 1996 Panel, questionnaire documentation is now available at the SIPP Web site at http://www.sipp.census.gov/sipp/.

3. The ranges for all legal values; and

4. The location (and size) in the file.

A machine-readable version of the data dictionary accompanies each data file. It can also be downloaded from the Internet (http://www.sipp.census.gov/sipp/).

The data dictionary is formatted to facilitate processing by user-written computer programs.[3] As shown in Figure 12-1, a "D" in the first column signifies that the next few lines define the variable: (1) the variable name, (2) the total number of columns occupied by the variable, (3) the starting position, (4) the number of occurrences of that variable, and (5) the size of each occurrence of the variable.[4] A "U" in the first column indicates that the next words describe the universe.[5] A "V" in the first column indicates that the next number and phrase describe one of the values of the variable. An asterisk in the first column denotes a comment. A period (.) before a word denotes the start of the value label.[6]

The format of the data dictionary for the longitudinal research files is different from that used for the core wave and topical module files. The full panel data dictionary includes two extra fields on the line with a "D" in the first column. The first extra field contains the number of occurrences of the variable, and the second extra field contains the number of digits for each occurrence of the variable. These fields are needed because some variables in the longitudinal research file occur *x* times, depending on the number of waves, or *y* times, depending on the number of months in the panel.

HH-ADDID in Figure 12-1 is a monthly variable containing two digits (monthly because it occurs 36 times). PP-MIS is also a monthly variable, but its length is one digit. PP-INTVW appears once per wave (because it occurs nine times), and PP-ENTRY, PP-PNUM, SU-TOTPP, and PP-RCSEQ occur once for the entire panel.

Figure 12-2 shows sample SAS and FORTRAN syntax for reading the data described by the codebook fragment in Figure 12-1. Additional SAS program code could be used to associate variable labels and value labels (SAS "formats") with the PP-MIS and PP-INTVW variables.

[3] The data dictionaries for the longitudinal research files use a different format from that used for the core wave and topical module files. Users who have worked with the core wave and topical module files should take care to note those differences. In addition, the formats of the data dictionaries for the 1996 Panel core wave and topical module files, as well as the variable names used in those files, have changed in the 1996 Panel. This chapter uses variable names from the 1990–1993 SIPP Panels. When longitudinal research files are released from the 1996 Panel, a revised version of this chapter will be available with updated information. Users will be able to download that version from the SIPP Web site at http://www.sipp.census.gov/sipp/.

[4] The data dictionary for the 1992 longitudinal research file used a different format from that used in the other pre-1996 longitudinal research files. In the 1992 data dictionary, the first line for each new variable, labeled with a "D" in column 1, has the following fields: variable name, total size (number of characters), start location, the length of a single occurrence of the variable, the number of occurrences of the variable, and the number of implied decimals.

[5] The universe definitions included in the data dictionaries prior to the 1996 Panel were often inaccurate. Users of pre-1996 SIPP Panels should check the skip patterns in the actual survey questionnaire to determine which subset of respondents was asked each question.

[6] The data dictionary for the 1992 longitudinal research file also has a line labeled with an "R" in column 1. This line provides the range of values for the variable.

Figure 12-1. Excerpt from the 1993 Longitudinal Research File Data Dictionary

```
D PP-ENTRY     2     17     1     2
    Range = (11:99)
    Edited entry address ID
    Address ID of the household that this person belonged to at the time this
      person first became part of the sample

D PP-PNUM      3     19     1     3
    Range = (101:999)
    Edited person number

D SU-TOTPP     2     22     1     2
    Range = (1:60)
    Total number of person records for this sample unit

D PP-RCSEQ     2     24     1     2
    Range = (1:60)
    Sequence number of person record within sample unit

D HH-ADDID     72     26     36     2
    Range = (0:99)
    Address ID. —— This field identifies the household this person lived in
      this month

D PP-INTVW     9     98     9    1
    Range = (0:4)
    Person's interview status for the relevant interview
V          0 .Not applicable (children under .15), not in sample, nonmatch
V          1 .Interview (self)
V          2 .Interview (proxy)
V          3 .Noninterview - Type Z refusal
V          4 .Noninterview - Type Z other

D PP-MIS     36     107     36     1
    Range = (0:2)
    Person's interview status for this month
V          0 .Not matched or not in sample
V          1 .Interview
V          2 .Non-interview
```

Figure 12-2. Corresponding SAS and FORTRAN Syntax to Read in Data from the 1993 Longitudinal Research File Data Dictionary

SAS

```
Input
   @17   PP_ENTRY   2.
         PP_PNUM    3.
         SU_TOTPP   2.
         PP_RCSEQ   2.
         (ADDID1-ADDID36)    (2.)
         (INTVW1-INTVW9)     (1.)
         (PP_MIS1-PP_MIS36) (1.)
         ;
```

FORTRAN

```
       INTEGER*2 PP_ENTRY
       INTEGER*2 PP_PNUM
       INTEGER*1 SU_TOTPP
       INTEGER*1 PP_RCSEQ
       INTEGER*1 HH_ADDID(36)
       INTEGER*1 PP_INTVW(9)
       INTEGER*1 PP_MIS(36)

       READ(infile,1000) PP_ENTRY, PP_NUM, SU_TOTPP,
      $         PP_RCSEQ, HH_ADDID, PP_INTVW, PP_MIS
1000   FORMAT(T17, I2, I3, I2, I2, 36I2, 9I1, 36I1)
```

Relationship of the Longitudinal Research Data Files to the SIPP Survey Instrument

The data dictionaries for the longitudinal research files do not replicate the survey instruments. Analysts should keep a few things in mind when using the data:

- The variables on the longitudinal research files do not correspond one-to-one with the questionnaire items. The variables are listed in a different order, some are not included in the longitudinal research file at all, and some are created from a combination of other variables.
- The range of possible values of the variables does not always correspond one-to-one with the response categories shown on the survey instrument or in the data dictionary;
- The variable name may not readily indicate its meaning; and

- The complexity of the skip patterns may not be apparent just by looking at the data dictionary.[7]

To avoid potential problems and confusion, users should become familiar with the survey instrument before using the data. When working with the data, analysts should refer to both the survey instrument and the data dictionary.

Structure of the Longitudinal Research Files

The longitudinal research files contain one record for each person who was ever in the SIPP sample for that panel. Even if the person was in the sample for just 1 month, there will be a record for that person. There are records for children as well as for adults, and there are records for people who entered the sample after the first wave.

Within each record, the variables correspond to the information that was collected in the core interviews. While most of the core items are included in the longitudinal research files, some items are not, and not all of the constructed variables found on the core wave files are included on the longitudinal research files. In addition, no items from any of the topical modules are included on the longitudinal research files. When items from the core wave or topical module files are needed, those variables must be merged with data from the longitudinal research files. Chapter 13 provides a detailed discussion of merging SIPP files.

The longitudinal research file structure differs from that of the core wave files. The longitudinal research files contain just one record per person, while the core wave files contain one record per person per month. Because some attributes do not change over the course of the panel, those variables appear once on each record (e.g., rotation group, sample unit ID, person number, sex, race, and ethnic origin). Some questions were asked once during each wave, so they appear *x* times on each record, where *x* equals the number of waves for that panel (e.g., highest grade attended, and participation in school breakfast and lunch programs). Most of the core questions were asked for each month of the panel. They appear *y* times on each record, where *y* equals the number of months for that panel (e.g., current address ID, monthly interview status, relationship to the reference person, income, and program participation).

Table 12-1 shows that the 1992 Panel has 10 waves (or 40 months) of data. The 1993 Panel has nine waves (or 36 months) of data. Thus, the interview status variable (PP-MIS) appears 40 times in the 1992 longitudinal research file, and it appears 36 times in the 1993 longitudinal research file.

[7] See footnote 5.

Table 12-1. Summary of Panels, Waves, Reference Months, and Sample Sizes

Panel Year	Reference Months	Number of Waves	Number of Months	Wave 1 Eligible Households
1984	Jun. 83 – Jun. 86	9	36	20,897
1985	Oct. 84 – Jul. 87	8	32	14,306
1986	Oct. 85 – Mar. 88	7	28	12,425
1987	Oct. 86 – Apr. 89	7	28	12,527
1988	Oct. 87 – Dec. 89	6	24	12,725
1989	Oct. 88 – Dec. 89	3	There is no longitudinal research file for the 1989 SIPP.	
1990	Oct. 89 – Aug. 92	8	32	23,627
1991	Oct. 90 – Aug. 93	8	32	15,626
1992	Oct. 91 – Mar. 95	10	40	21,577
1993	Oct. 92 – Dec. 95	9	36	21,823

Source: SIPP Quality Profile, 3rd Ed. (U.S. Census Bureau, 1998a).

Table 12-2 illustrates the longitudinal research file structure. In this example, there are five people. Sample unit ID (PP-ID), person number (PP-PNUM), and entry address ID (PP-ENTRY) appear once on each record because they are permanent characteristics of those people. Monthly interview status (PP-MIS), a monthly variable, appears 40 times because the 1992 Panel had 10 waves and each wave collected information about the 4 months prior to the interview month.

People who were not interviewed (in person or by proxy) for 1 or more months over the course of the panel either have their data imputed[8] or are identified as not in the sample (PP-MIS equal to either 0 or 2) for the months when they were not in the sample. The discussion of the PP-MIS variable later in this chapter provides additional information.

How to Align Data by Calendar Month

It is frequently useful to realign the SIPP data by calendar month instead of reference month. For example, researchers often want to analyze data for a specific calendar year (January through December) or federal fiscal year (October through September).[9] To do this, the analyst must

[8] Imputation would be by Type Z and missing-wave imputations. Chapter 4 discusses imputation methods.

[9] The longitudinal research files do not contain calendar month weights. Those weights would be needed for some types of longitudinal analyses, such as analyses of the dynamics of program participation, where the unit of analysis is a spell of program participation (Chapter 8 provides a discussion of this example). Data from the longitudinal research files can also be used for cross-sectional estimation, and they are often preferable to the data from the core wave files because the edit and imputation procedures used for the longitudinal research files are believed to result in less imputation error than the procedures used for the core wave files. The format of the file is sometimes easier to work with, even for cross-sectional applications. In those instances, the calendar month weights must be merged from the core wave files. Chapter 8 provides a detailed discussion of weighting procedures in the SIPP. Chapter 13 provides a detailed discussion of linking SIPP files.

Table 12-2. Example of the Longitudinal Research File Structure

PP-ID	PP-ENTRY	PP-PNUM	PP-ROT	PP-MIS																			
				Wave 1				Wave 2				Wave 3				Wave 4				Wave 5			
				Month				Month				Month				Month				Month			
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
112612345	11	101	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
112987122	11	101	2	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
987913389	11	101	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
123912879	11	101	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
123912879	11	201	3	0	0	0	0	0	1	1	1	1	1	1	1	2	2	1	1	1	1	1	0
874943283	11	101	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	101	4	1	1	1	0	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
788723892	11	102	4	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	0	0	0	0
788723892	11	301	4	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	1001	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
763483873	11	101	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
890987123	11	101	1	1	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1	1	1	2
				PP-MIS																			
				Wave 6				Wave 7				Wave 8				Wave 9				Wave 10			
				Month				Month				Month				Month				Month			
PP-ID	PP-ENTRY	PP-PNUM	PP-ROT	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
112612345	11	101	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
112987122	11	101	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
987913389	11	101	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
123912879	11	101	3	2	1	1	1	0	0	2	2	2	0	0	0	0	0	0	0	0	0	0	0
123912879	11	201	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
874943283	11	101	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	101	4	1	1	1	1	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
788723892	11	102	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
788723892	11	301	4	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
788723892	11	1001	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
763483873	11	101	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
890987123	11	101	1	2	2	1	1	1	1	1	1	1	1	2	2	2	1	1	1	0	0	0	0

know the reference period for each rotation group of the panel. That information is included with the technical documentation that accompanies the longitudinal research files.

Table 12-3 shows the reference period for each rotation group of the 1992 Panel. It shows that the reference period for rotation group 2 is October 1991–January 1995. The reference period for rotation group 3 is November 1991–February 1995. The reference period for rotation group 4 is December 1991–March 1995. The reference period for rotation group 1 is January 1992–December 1994 (interviews were not conducted in Wave 10 for this rotation group).

Table 12-3. Reference Periods for Each Rotation Group of the 1992 Panel

Rotation Group (ROT)	Reference Period
2	October 1991–January 1995
3	November 1991–February 1995
4	December 1991–March 1995
1	January 1992–December 1994

The following algorithm (Figure 12-3), written for the 1992 Panel, illustrates one approach to realigning the SIPP reference months to common calendar months. The mapping depends on the panel and rotation group and must be applied to each person. The first step establishes the displacement or realignment of the months. The second step initializes each monthly variable to –9 to distinguish the calendar months in which the variable is not relevant.[10] The loop goes from 1 to 42 because in the 1992 Panel the first reference month was October 1991 and the last reference month was March 1995, which means that there were 42 calendar months covered by the panel. The third part of the algorithm realigns the input data to be based on the calendar month. Table 12-4 displays the data after the realignment.

Using the Monthly Interview Status (PP-MIS) Variables

The monthly interview status variable helps to determine whether the data for a person in a given month should be used. In the longitudinal research files, this variable is labeled PP-MIS, and it has one occurrence for each reference month of the SIPP panel. Some people refer to it as the *in-sample* variable to distinguish it from the interview status variable (PP-INTVW). The PP-MIS variables have three possible values: 0, 1, and 2.

[10] If –9 is a possible value for the variables being realigned (e.g., self-employed income can be negative), a different starting value must be used.

Figure 12-3. Algorithm for Realigning SIPP Panel Month to Calendar Months in the 1992 Panel

```
/*
Create a variable that identifies the number of months each
rotation group differs from the baseline
*/
If ROT = 2
   DISPLACEMENT = 0
Else if ROT = 3
   DISPLACEMENT = 1
Else if ROT = 4
   DISPLACEMENT = 2
Else if ROT = 1
    DISPLACEMENT = 3
End if

/*
Initialize the new, re-aligned variable. This is not needed in SAS.
When this step is used, an initial value should be chosen that
is not a legal value for the variable in the actual data.
*/
For each calendar month (for CALMM = 1 to 42):
   NEW-PP-MIS(CALMM) = -9
End loop

/*
Create the newly re-aligned variable
*/
For each reference month (for MONTH = 1 to 40):
   CALMM = MONTH + DISPLACEMENT
   NEW-PP-MIS(CALMM) = PP-MIS(MONTH)
End loop
```

The monthly interview status is the only reliable guide to whether the data for a given person should be used in a given month. Analysts should use only data for those months in which a person's interview status (PP-MIS) is equal to 1.[11]

Any data present for months in which a person's interview status is coded either 0 or 2 should be ignored. A code of 0 indicates that the person was not in the sample that month, and a code of 2 indicates a noninterview for that month.[12]

[11] As a safeguard against inadvertently using data for months when PP-MIS is not equal to 1, all monthly variables in the user's data extract should be set to a missing value for months when PP-MIS is not equal to 1. Most statistical packages allow certain values to be flagged as "missing." Once flagged, those values are excluded from computations.

[12] Beginning with the 1991 Panel, new "missing wave" imputation procedures were instituted for the longitudinal research files. Whenever data for a wave are imputed (the WAVFLG variable), PP-MIS is recoded to 1 on the longitudinal research files, indicating that the data for those months should be used. In some cases, these people will have records in the core wave files that were created during the Type Z imputation processing (see Chapter 4 for details). In some of these instances, however, the longitudinal research file will have data for people who are not present on the associated core wave data files.

Table 12-4. Monthly Data from the 1992 Panel, Realigned by Calendar Month

				NEW-PP-MIS													
				1991			1992										
PP-ID	PP-ENTRY	PP-PNUM	PP-ROT	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
112612345	11	101	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
112987122	11	101	2	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
987913389	11	101	3	-9	1	1	1	1	1	1	1	1	1	1	1	1	1	1
123912879	11	101	3	-9	1	1	1	1	1	1	1	1	1	1	1	1	1	1
123912879	11	201	3	-9	0	0	0	0	0	1	1	1	1	1	1	1	2	2
874943283	11	101	4	-9	-9	1	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	101	4	-9	-9	1	1	1	0	0	1	1	1	1	1	1	1	0
788723892	11	102	4	-9	-9	1	1	1	1	1	1	1	1	1	1	1	1	2
788723892	11	301	4	-9	-9	0	0	0	0	1	1	1	1	1	1	1	1	1
788723892	11	1001	4	-9	-9	0	0	0	0	0	0	0	0	0	0	0	0	0
763483873	11	101	1	-9	-9	-9	1	1	1	1	1	1	1	1	1	1	1	1
890987123	11	101	1	-9	-9	-9	1	1	1	1	1	1	1	1	1	2	2	2

				NEW-PP-MIS											
				1993											
PP-ID	PP-ENTRY	PP-PNUM	PP-ROT	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
112612345	11	101	2	1	1	1	1	1	1	1	1	1	1	1	1
112987122	11	101	2	0	0	0	0	0	0	0	0	0	0	0	0
987913389	11	101	3	1	1	1	1	1	1	1	1	1	1	1	1
123912879	11	101	3	1	1	1	1	1	2	2	1	1	1	0	0
123912879	11	201	3	1	1	1	1	1	0	0	0	0	0	0	0
874943283	11	101	4	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	101	4	0	1	1	1	1	1	1	1	1	1	1	2
788723892	11	102	4	2	2	2	0	0	0	0	0	0	0	0	0
788723892	11	301	4	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	1001	4	0	0	0	0	0	0	0	0	0	0	0	0
763483873	11	101	1	1	1	1	1	1	1	1	1	1	1	1	1
890987123	11	101	1	1	1	1	1	1	1	1	2	2	2	1	1

(table continues)

Table 12-4. Monthly Data from the 1992 Panel, Realigned by Calendar Month (*continued*)

				NEW-PP-MIS														
				1994												1995		
PP-ID	PP-ENTRY	*PP-PNUM*	PP-ROT	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
112612345	11	101	2	1	1	1	1	1	1	1	1	1	1	1	1	1	–9	–9
112987122	11	101	2	0	0	0	0	0	0	0	0	0	0	0	0	0	–9	–9
987913389	11	101	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	–9
123912879	11	101	3	2	2	2	0	0	0	0	0	0	0	0	0	0	0	–9
123912879	11	201	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	–9
874943283	11	101	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	101	4	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
788723892	11	102	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
788723892	11	301	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
788723892	11	1001	4	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
763483873	11	101	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
890987123	11	101	1	1	1	1	1	1	1	2	2	2	1	1	1	0	0	0

The presence of data in analysis fields for any given month is not a reliable guide to whether the person should be included in the planned analyses. Data are collected for all months of the reference period for a given wave, even if the interviewed person was in the sample for only part of the reference period. Data are also present even if the person was not interviewed. Information from the questionnaire is imputed when the person was in sample for at least 1 month of the reference period but not actually interviewed. That includes people who moved out of scope (as defined in Chapter 2), people who died, and people who refused to be interviewed. The entire questionnaire was imputed for Type Z noninterviews (people who refused to be interviewed, living in households where other members were successfully interviewed). Chapter 4 examines imputation procedures; Chapter 8 provides information on weighting. Data are collected for all months of the reference period even if the interviewed person was in the sample for only part of the reference period.

The presence of a positive weight is also not a reliable guide to whether a person should be included in the planned analysis. Although people with zero weights will not enter into any weighted tabulations, they may provide important contextual information about people who do enter into those (weighted) tabulations. For example, a zero-weight person who is a member of the same household as a positive-weight person for only 3 months provides information about the positive-weighted person's household (including, for example, household size, composition, income, and program participation) for that 3-month period. That is why records for these zero-weighted people are retained in the SIPP full panel data files.[13]

Identifying Persons

There are many occasions when a user may need to identify which records belong to each individual in the SIPP data files. That need arises, for example, during the following procedures:

- Merging data from topical module or full panel files to core wave files;
- Combining data from two or more core wave files;
- Linking husbands and wives;
- Linking parents and children; and
- Identifying which person received government transfer income on behalf of the family.

To uniquely identify a person in the longitudinal research files, analysts should use the three variables shown in Table 12-5.[14]

[13] Using the PP-MIS variable shown in Table 12-2, one can see that the first person within each rotation group was in sample every month of the panel. The second person shown in the table left the sample before the third interview (information was probably collected by proxy interview for that wave) and did not return to the sample. The eighth person left the sample in month 13. The tenth person entered the sample in month 38 (the last wave).

[14] Beginning with the 1996 Panel, the entry address ID will no longer be needed: person numbers will be unique within sample units. Continued use of the entry address ID will not create any problems. It is simply redundant information.

Table 12-5. Variables Used to Uniquely Identify a Person in the Longitudinal Research Files

Variable Name	Description
PP-ID	Sample unit ID
PP-ENTRY	Entry address ID
PP-PNUM	Person number

- PP-ID uniquely identifies each initially sampled dwelling unit.[15] Every person in the longitudinal research file was either a member of one of those units (an original sample member) or lived with someone during the life of the panel who was a member of an initially sampled dwelling unit. A person's connection to that unit is an attribute of that person and does not change over time.[16] This means that as people move from address to address, their PP-ID stays the same. As new people join the homes of original sample members, they receive the PP-ID of the original sample members.
- PP-ENTRY identifies the address where the person lived at the time he or she was first interviewed. It does not change even if the person moves.[17] It is used in conjunction with the person number and the sample unit ID to uniquely identify persons within the sampling unit. Values for this variable are unique only within sample units. The entry address ID has two components. The first part of the ID number (two digits in the 1992 Panel, and one digit in all others) identifies the wave in which SIPP interviews were first conducted at the address. The second part of the number (one digit in all panels) sequentially numbers addresses within a sample unit (PP-ID) that enter the sample in the same wave.
- PP-PNUM uniquely identifies a person within the sample unit ID and entry address ID. PP-PNUM does not change even if the person moves.[18] The first part of PP-PNUM (two digits in the 1992 Panel, and one digit in all others) indicates the wave in which the person was first interviewed.[19] The remaining two digits are sequentially assigned within the household. Thus, original sample members are assigned person numbers ranging from 100 to 199. Individuals who enter the SIPP sample in Wave 2 are assigned a person number ranging from 200 to 299. Those who enter in Wave 10 are assigned person numbers ranging from 1001 to 1099.

Table 12-6 illustrates how the combination of PP-ID, PP-ENTRY, and PP-PNUM uniquely identifies people and provides information about when they first entered the SIPP sample. In this example, there are eight individuals: five are original sample members; one person joined the

[15] The PP-ID is a random recode of three other variables in the Census Bureau's internal (not public use) files: the respondent's sampling area (PSU), the cluster of housing units within that area (called the "segment"), and a sequentially assigned serial number. Those three variables are omitted from the public use files to protect the confidentiality of the respondents.

[16] There is one rare exception to this rule, which is described in the section entitled "Identifying Movers" later in this chapter.

[17] See footnote 16.

[18] See footnote 16.

[19] For Wave 10 of the 1992 Panel and for the 1996 Panel, the first two digits of PNUM instead of the first digit identify the wave in which the person entered the sample.

SIPP sample in Wave 4, one person joined in Wave 7, and one person joined in Wave 10 (of the 1992 Panel).

Table 12-6. How to Uniquely Identify a Person in the Longitudinal Research Files

Sample Unit ID (PP-ID)	Entry Address ID (PP-ENTRY)	Person Number (PP-PNUM)	Notes
123456789	11	101	Original sample member
123456789	11	102	Original sample member
123456789	11	401	Enters SIPP sample in Wave 4
123456789	71	701	Enters SIPP sample in Wave 7
321456789	11	101	Original sample member
321456789	11	102	Original sample member
321456789	11	103	Original sample member
456789123	101	1001	Enters SIPP sample in Wave 10 of the 1992 Panel

Identifying Households

The term *household*, as used in Census Bureau publications, refers to a group of people who occupy a housing unit. A house, an apartment or other group of rooms, or a single room is regarded as a housing unit if it is occupied or intended for occupancy as separate living quarters. That is, the occupants do not live and eat with any other people in the structure and there is direct access from the outside or through a common hall. A group of friends sharing an apartment constitutes a household. Rooming and boarding houses, college dormitories, convents, and monasteries are classified as *group quarters* rather than households.

To uniquely identify a household or group quarters in the longitudinal research files in a given month, analysts should use the variables shown in Table 12-7.[20]

Table 12-7. Variables Used to Uniquely Identify a Household in the Longitudinal Research Files

Variable Name	Description
PP-ID	Sample unit ID
HH-ADDID$_i$	Current address ID in the *i*th month
PP-MIS$_i$	Person's interview status in the *i*th month

[20] Since household composition changes from one month to the next, it is generally not possible to construct "longitudinal households." Users should not infer commonality across months based solely on place of residence in one month. The characteristics of the household to which a given person belongs (such as household size and household income) should be evaluated separately for each month, based on just those people who reside together in each specific month. Similar caution should be exercised when dealing with the characteristics of the family and, when applicable, the subfamily to which a person belongs.

People with the same PP-ID and HH-ADDID$_i$ values and with a PP-MIS value of 1 live in the same household (or group quarters) in the *i*th month of the reference period. The eight individuals shown in Table 12-8 make up four households. The first household contains the first four individuals. The second household contains one person. The third household contains one person. The fourth household contains two people.

This example depicts the households in the *i*th month. These people could belong to different households in other months. (Users may find it helpful when reading the following pages to refer to Figure 2-1 [pp. 2-10–2-14], which illustrates changes in household composition.)

Table 12-8. How to Uniquely Identify a Household or Group Quarters in a Given Month of the Longitudinal Research Files

Sample Unit ID (PP-ID)	Entry Address ID (PP-ENTRY)	Person Number (PNUM)	Person's Interview Status (PP-MIS)	Current Address ID (HH-ADDID$_i$)	Notes
123456789	11	101	1	71	Four people in this household
123456789	11	102	1	71	
123456789	11	401	1	71	
123456789	71	701	1	71	
321456789	11	101	1	31	One person in this household
321456789	11	102	1	32	One person in this household
321456789	11	103	1	101	Two people in this household [a]
321456789	101	1001	1	101	

[a] Because this example includes a person with an entry address of 101, we know that the example refers to a month from Wave 10 of the 1992 Panel (the only panel prior to 1996 with 10 or more waves).

Identifying Families

The term *family*, as used in Census Bureau publications, refers to a group of two or more people related by birth, marriage, or adoption who reside together; all such individuals are considered members of one family.[21]

- A *primary family* is a family containing the household reference person and all of his or her relatives. This means that a household composed of a husband and wife, their son, and their son's wife (i.e., the daughter-in-law) is classified as a primary family containing four people.

[21] As with households (see footnote 20), because family composition changes from one month to the next, it generally is not possible to construct longitudinal families. Users should not infer commonality across months based solely on family membership in one month. The characteristics of the family to which a person belongs (such as family size and family income) should be evaluated separately for each month, and should be based on just those people who reside together and are members of the same family in each specific month. Similar caution should be exercised when dealing with the characteristics of the household and, when applicable, the subfamily (related or unrelated) to which a person belongs.

- A *related subfamily* is a nuclear family that is related to but does not include the household reference person. For example, the son and his wife (i.e., the daughter-in-law) in the preceding example are a related subfamily.
- An *unrelated subfamily* (sometimes called a secondary family) is a nuclear family that is not related to the *household* reference person. Thus, a husband and wife who live in a friend's house are classified as an unrelated subfamily. A mother and daughter who live in the mother's boyfriend's apartment are classified as an unrelated subfamily.
- A *primary individual* is a household reference person who lives alone or lives with only nonrelatives. Primary individuals are sometimes treated by the Census Bureau as families with only one person and are referred to as pseudo-families.
- A *secondary individual* is not a household reference person and is not related to any other people in the household. Secondary individuals are sometimes treated by the Census Bureau as families with only one person and are referred to as pseudo-families.

Unlike the core wave files, the longitudinal research files do not contain family identification variables (e.g., FID, FID2, and SID). Analysts needing family identification variables must either merge them from the core wave files (Chapters 10 and 13) or create them.[22] Because family composition can change over time, these are monthly variables. The algorithm in Figure 12-4 shows one approach to creating functional equivalents of the variables contained on the core wave files.[23]

The variables created by this algorithm are functionally equivalent to the variables with the same names on the core wave files: they will group people into the same family and subfamily groups. However, the actual values assigned by this algorithm to these variables generally will not equal the values found in the variables from the core wave files.

With these monthly variables (FID_i, $FID2_i$, and SID_i), users can identify common family membership in each month.[24] The Census Bureau has two principal methods for distinguishing families that are based on the variables and numbering schemes shown in Table 12-9. Analysts must remember to choose which type of family classification they want and then use the appropriate method.

- The first method defines a family as all persons who are related and living together. The family ID variable FID_i is used with this definition. FID_i groups the household reference person with all related household members by assigning them the same ID number.

[22] In most cases, it is also possible to merge these variables from the core wave files. However, beginning with the 1991 Panel, a missing wave imputation procedure was applied to the longitudinal research files: data were imputed for people with missing data for a wave but with valid data for the two adjacent waves. Although these people have data in the longitudinal research file for imputed waves, some have no data in the core wave files (some of these people are subject to Type Z imputation procedures that create records in the core wave files). For these people, merging the family ID variables from the core wave files is not an option.

[23] This algorithm uses the following (monthly) variables found on the longitudinal research files: FAMTYP and FAMNUM. These variables are discussed in greater detail in the next section.

[24] See footnotes 20 and 21.

Figure 12-4. Constructing Family and Subfamily ID Variables in the Longitudinal Research Files

```
For each person (index = ip):
      For each month (index = mo):
            If PP-MIS(mo, ip)= 1 then do:       <i.e., interview status>
                  If FAMTYP(mo, ip) = 0         <i.e., primary family>
                     then FID(mo, ip)  = 1
                        FID2(mo, ip) = 1
                        SID(mo, ip) = 0
                  Else if FAMTYP(mo, ip) = 1    <i.e., secondary individual>
                     then FID(mo, ip) = 10000 + ip
                        FID2(mo, ip) = 10000 + ip
                        SID(mo, ip) = 0
                  Else if FAMTYP(mo, ip) = 2    <i.e., unrelated subfamily>
                     then FID(mo, ip) = 100 + FAMNUM(mo, ip)
                        FID2(mo, ip) = 100 + FAMNUM(mo, ip)
                        SID(mo, ip) = 0
                  Else if FAMTYP(mo, ip) = 3    <i.e., related subfamily>
                     then FID(mo, ip) = 1
                        FID2(mo, ip) = 0
                        SID(mo, ip) = FAMNUM(mo, ip)
                  Else if FAMTYP(mo, ip) = 4     <i.e., primary individual>
                     then FID(mo, ip) = 10000 + ip
                        FID2(mo, ip) = 10000 + ip
                        SID(mo, ip) = 0
                  End if
            End "PP-MIS = 1" Block
      End month loop
End person loop
```

Table 12-9. Variables Used to Identify Families in the Longitudinal Research Files

Variable Name	Description
PP-ID	Sample unit ID
HH-ADDID$_i$	Address ID in the *i*th month
PP-MIS$_i$	Person's interview status in the *i*th month
And one of the following created variables:	
FID$_i$	Family ID in the *i*th month
FID2$_i$	Family ID in the *i*th month, excluding related subfamily members (FID2$_i$ equals zero for related subfamily members)
SID$_i$	Family ID in the *i*th month for related subfamily members (SID$_i$ assigns nonzero values only to members of related subfamilies)
FID2$_i$ and SID$_i$	Family ID in the *i*th month, separating related subfamilies from the primary family

Note: Variables FID$_i$, FID2$_i$, and SID$_i$ are not included on the longitudinal research files. They can be created by using the algorithm shown in Figure 12-4 or merged from the core wave files.

This family group corresponds to the Census Bureau's definition of a primary family. FID_i groups members of each unrelated subfamily (and primary and secondary individuals) separately.

- The second method is similar to the first in defining a family, but the family excludes related subfamilies. The family ID variable $FID2_i$ is used with this definition. $FID2_i$ equals zero for related subfamilies.

Analysts who want to analyze multigenerational families would use $FID2_i$ and the variable SID_i. SID_i treats related subfamilies as distinct family units by assigning them nonzero values. Analysts can easily distinguish unrelated subfamilies from other family units when they use these variables and numbering schemes.

Table 12-10 illustrates the difference between FID_i, $FID2_i$, and SID_i for a single month. In the month shown, the first household contains a primary family of five people. The primary family contains two related subfamilies. FID_i and $FID2_i$ mask the fact that there are two related subfamilies; only SID_i provides that information. SID_i has nonzero values only for members of related subfamilies. The second household contains a primary family and two unrelated subfamilies. The third household contains a primary individual and an unrelated subfamily. The fourth household contains only a primary individual. The fifth household is group quarters containing two people. This example depicts those families in the *i*th month. These people could belong to different families in other months.[25]

The specific analysis being planned will inform the choice of which family classification to use. To group people into families in the same way that the Census Bureau does, analysts should use PP-ID, $PP\text{-}MIS_i$, $HH\text{-}ADDID_i$, and FID_i. To analyze primary families excluding related subfamily members, analysts should include only those records with $FID2_i$ greater than zero. To analyze related subfamilies as distinct family units, analysts should use only those records with SID_i greater than zero. To uniquely identify (1) primary families excluding related subfamilies and (2) related subfamilies treated as distinct family groups, analysts should use PP-ID, $PP\text{-}MIS_i$, $HH\text{-}ADDID_i$, $FID2_i$, and SID_i. In those analyses, it is easy to distinguish unrelated families from other families.

Variables Describing Household and Family Composition

Table 12-11 shows the variables contained on the longitudinal research files summarizing household and family composition.[26]

[25] See footnote 18.

[26] More detailed information about the relationships between members is collected in the Household Relationships topical module. Those data provide extensive information about household composition at the time of the topical module interview.

Table 12-10. How to Uniquely Identify a Family in a Given Month of the Longitudinal Research Files

Sample Unit ID (PP-ID)	Current Address ID (HH-ADDID$_i$)	Person's Interview Status (PP-MIS$_i$)	Family ID, Including Subfamily (FID$_i$)	Family ID, Excluding Subfamily (FID2$_i$)	Subfamily ID (SID$_i$)	Family Type (FAMTYP$_i$)	Person Number (PP-PNUM)	Notes
110011111	11	1	1	1	0	0	101	This household contains a primary family of five people. The primary family contains two related subfamilies.
110011111	11	1	1	0	2	3	102	
110011111	11	1	1	0	2	3	103	
110011111	11	1	1	0	3	3	104	
110011111	11	1	1	0	3	3	105	
122210000	33	1	1	1	0	0	101	This household contains a primary family and two unrelated subfamilies.
122210000	33	1	1	1	0	0	104	
122210000	33	1	101	101	0	2	305	
122210000	33	1	101	101	0	2	306	
122210000	33	1	102	102	0	2	307	
122210000	33	1	102	102	0	2	308	
555555555	21	1	1001	1001	0	4	101	This household contains a primary individual and an unrelated subfamily.
555555555	21	1	101	101	0	2	201	
555555555	21	1	101	101	0	2	202	
555555555	21	1	101	101	0	2	203	
610000000	11	1	1001	1001	0	4	101	Primary individual.
897454644	11	1	1001	1001	0	1	101	Group quarters with two secondary individuals.
897454644	11	1	1002	1002	0	1	102	

Notes: Variables FID$_i$, FID2$_i$, and SID$_i$ are not part of the longitudinal research files. They can be merged from the core wave files or created using the algorithm shown in Figure 12-4. FAMTYP = 0 means the person belongs to a primary family. FAMTYP = 1 means the person is a secondary individual. FAMTYP = 2 means the person belongs to an unrelated subfamily. FAMTYP = 3 means the person belongs to a related subfamily. FAMTYP = 4 means the person is a primary individual.

Table 12-11. Variables Used to Describe Household Composition in the Longitudinal Research Files

Variable Name	Description
$FAMTYP_i$	Type of family in the *i*th month (e.g., primary family, related subfamily)
$FAMREL_i$	Family relationship in the *i*th month (e.g., reference person, spouse of family reference person, child of family reference person)
RRP_i	Recoded relationship to the household reference person in the *i*th month (e.g., household reference person living with relatives, child of household reference person)
$ENTID\text{-}SP_i$	Entry address ID of spouse in the *i*th month
$PNSP_i$	Person number of spouse in the *i*th month
$ENTID\text{-}PT_i$	Entry address ID of parent in the *i*th month
$PNPT_i$	Person number of parent in the *i*th month
$U\text{-}PNG_j$	Person number of guardian in the *j*th wave
$ENTID\text{-}GD_j$	Entry address ID of guardian in the *j*th wave

As Table 12-12 shows, RRP_i summarizes the relationship of each person to the *household reference person* in month *i*.

Table 12-12. Relationship to the Household Reference Person in a Given Month

Edited Relationship to the Household Reference Person (RRP_i)	Description
1	Household reference person, living with relatives
2	Household reference person, living alone or with nonrelatives
3	Spouse of household reference person
4	Child of household reference person
5	Other relative of household reference person
6	Nonrelative of household reference person, but related to other members of the household
7	Nonrelative of all members of the household

The household description depends on the identity of the reference person. For example, in Table 12-13, the household contains a mother, her daughter, and her daughter's son. If the mother is the household reference person ($RRP_i = 1$), her daughter is listed as a child of the household reference person ($RRP_i = 4$) and the daughter's son is listed as other relative of the household reference person ($RRP_i = 5$). If the daughter is the reference person, her son is listed as a child of the household reference person ($RRP_i = 4$) and her mother is listed as other relative of the household reference person ($RRP_i = 5$). Users should note that the household reference person can change from one month to the next; thus, the household description could also change.

Table 12-13. Using RRP to Identify Households Containing Three Generations in the Longitudinal Research Files

Household Reference Person	Relationship to the Household Reference Person (RRP_i)	Notes
Mother as Household Reference Person		
Mother	1	Reference person
Daughter	4	Child of reference person
Daughter's son	5	Other relative of reference person
Daughter as Household Reference Person		
Daughter	1	Reference person
Daughter's son	4	Child of reference person
Mother	5	Other relative of reference person

Six other variables in the longitudinal research file can be used to describe household and family composition: $PNSP_i$, ENTID-SP_i, $PNPT_i$, ENTID-PT_i, U-PNG_j, and ENTID-GD_j. These six variables identify the person number and entry address ID of the spouse, parent, or guardian living at the same address as the person in the *i*th month or *j*th wave (in the last two cases).[27] By building from these variables, the analyst can identify a variety of family configurations. For example, these variables can be used to identify households containing three generations. Table 12-14 displays one household containing a mother and her two children. One child (PP-PNUM = 102) has a son, and the other child (PP-PNUM = 104) has a spouse.

Table 12-14. Using PNSP and PNPT to Identify Households Containing Three Generations in the Longitudinal Research Files

Household Member	Entry Address ID (PP-ENTRY)	Person Number (PP-PNUM)	Relationship to Household Reference Person (RRP_i)	Entry Address ID of Spouse (ENTID-SP_i)	Spouse ($PNSP_i$)	Entry Address ID of Parent (ENTID-PT_i)	Parent ($PNPT_i$)	Notes
Mother	11	101	1	11	999	11	999	Mother
Daughter #1	11	102	4	11	999	11	101	Child
Daughter #1's son	11	103	5	11	999	11	102	Grandchild
Daughter #2	11	104	4	11	105	11	101	Child
Spouse of Daughter #2	11	105	5	11	104	11	999	Spouse of child

Note: Value of 999 means not applicable.

[27] Parents and spouses always share the same sample unit ID (PP-ID) as the respondent. The variables are assigned values only in the months that people are living together. For example, a couple living together in Wave 1 would have values in the PNSP and ENTID-SP variables that pointed to each other. However, if they separate (and remain married) in Wave 2, the PNSP and ENTID-SP variables will be assigned values of 999 (indicating that the variables are not applicable).

Using Family-Level Income Variables

The longitudinal research files contain a number of family-level income variables. The family income variables on the longitudinal research files include the income of all related subfamily members. *In other words, primary family members and related subfamily members are treated as one family by the Census Bureau when calculating family-level income amounts*. The longitudinal research files do not contain any subfamily income variables. If family income variables are needed that do *not* pool related subfamilies with primary families, those income variables must be created. That is done by looping over persons with PP-MIS$_i$ of 1 and with common PP-ID, HH-ADDID$_i$, FID2$_i$, and SID$_i$ for each month.[28]

Table 12-15 illustrates how the family income variables on the longitudinal research files include the income of related subfamily members. From the previous example of a primary family of five people, the primary family contains two related subfamilies. Total family income (FF-INC$_i$) is $3,100. The incomes of all subfamily members are included in that amount.

Table 12-15. Family Income in the Longitudinal Research Files

Sample Unit ID (PP-ID)	Entry Address ID (PP-ENTRY)	Person Number (PP-PNUM)	Person Interview Status (PP-MIS$_i$)	Current Address ID (HH-ADDID$_i$)	Family ID, Including Subfamily (FID$_i$)	Sub-family ID (SID$_i$)	Total Family Income (FF-INC$_i$)	Person-Level Income (PP-INC$_i$)
110011111	11	101	1	11	1	0	$3,100	$ 100
110011111	11	102	1	11	1	2	$3,100	$ 500
110011111	11	103	1	11	1	2	$3,100	$ 500
110011111	11	104	1	11	1	3	$3,100	$1,000
110011111	11	105	1	11	1	3	$3,100	$1,000

More About Using the SIPP ID Variables: Identifying Movers

When a person moves, the current address field (HH-ADDID$_i$) changes. The PP-ID, PP-ENTRY, and PP-PNUM values remain the same. The first digit (or first two digits in the 1992 Panel) of HH-ADDID$_i$ indicate(s) the wave in which a household is first interviewed at that new address. The remaining digits sequentially number the households that split into two or more households, as a result of a move to a different location by original sample members. Thus, new addresses in Wave 2 are numbered 21, 22, and so on. New addresses in Wave 3 are numbered 31, 32, and so on. New addresses in Wave 10 are numbered 101, 102, and so on. (Readers may wish to refer to Figure 2-1 [pp. 2-10–2-14], which illustrates movement into and out of households.)

[28] FID$_i$ and SID$_i$ are not included on the longitudinal research files. They can be merged from the core wave files or created by using the algorithm shown in Figure 12-4.

Table 12-16 shows that persons 101 and 102 in the first household are original sample members. Person 401 moved into the home of persons 101 and 102 in Wave 4. In Wave 7, all three moved to a new location and were joined by person 701. In the second household, person 101 is an original sample member who moved to a new location in Wave 3. In the third household, person 102 is an original sample member who used to live with persons 101 and 103 of the same sample unit ID (PP-ID), but moved to a new location in Wave 3 (to a different location from person 101). In the fourth household, person number 103 is an original sample member who used to live with persons 101 and 102 of the same sample unit ID number. Person 103 moved to a new location in Wave 10 and was joined by person 1001, who just entered the SIPP sample. All but two people moved from their original location (i.e., only two people have HH-ADDID$_i$ equal to PP-ENTRY).

Table 12-16. How to Identify Movers in the Longitudinal Research Files

Wave	Sample Unit ID (PP-ID)	Entry Address ID (PP-ENTRY)	Person Number (PP-PNUM)	Person Interview Status (PP-MIS$_i$)	Current Address ID (HH-ADDID$_i$)	Notes
1	123456789	11	101	1	11	Persons 101 and 102 are the original sample members
	123456789	11	102	1	11	
4	123456789	11	101	1	11	Person 401 begins to live with them in Wave 4.
	123456789	11	102	1	11	
	123456789	11	401		11	
7	123456789	11	101	1	71	All three people move in Wave 7 and person 701 joins them
	123456789	11	102	1	71	
	123456789	11	401	1	71	
	123456789	71	701		71	
1	321456789	11	101	1	11	Person 101, person 102, and person 103 are original sample members.
	321456789	11	102	1	11	
	321456789	11	103	1	11	
3	321456789	11	101	1	31	Person 101 moved in Wave 3. Person 102 moved in Wave 3 to a different location from person 101. Person 103 remained with person 101.
	321456789	11	102	1	32	
	321456789	11	103	1	31	
10	321456789	11	101	1	31	Person 103 is an original sample member who used to live with persons 101 and 102 of the same ID. In Wave 10, person 103 lives in a new location with person 1001, who just entered the SIPP sample.
	321456789	11	102	1	32	
	321456789	11	103	1	101	
	321456789	101	1001	1	101	

The next example (Table 12-17) further illustrates how the ID system works as people move to new addresses, additional people move in with them, and households split. A review of Figure 2-1 (pp. 2-10–2-14) may help in understanding the various household changes.

- In Wave 1, there is a five-person household consisting of a husband, a wife, a daughter, a son, and a cousin. Because this is the first wave, the current address number is 11, indicating

Table 12-17. Another Example of Household Changes and Their Effects on the ID Variables in the Longitudinal Research Files

Household Member	Sample Unit ID (PP-ID)	Current Address ID (HH-ADDID$_i$)	Entry Address ID (PP-ENTRY)	Person Number (PP-PNUM)
Wave 1				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son	101111103	11	11	104
Cousin	101111103	11	11	105
Wave 2				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son	101111103	11	11	104
Cousin	101111103	11	11	105
Wave 3				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son-in-Law	101111103	11	11	301
Cousin	101111103	11	11	105
Wave 4	**Parent's Household**			
Father	101111103	11	11	101
Mother	101111103	11	11	102
	Daughter's Household			
Daughter	101111103	41	11	103
Son-in-Law	101111103	41	11	301
	Cousin's Household			
Cousin	101111103	42	11	105
Uncle	101111103	42	42	401
Wave 10	**Parent's Household**			
Father	101111103	11	11	101
Mother	101111103	11	11	102
	Daughter's Household			
Daughter	101111103	41	11	103
Son-in-Law	101111103	41	11	301
Newborn	101111103	41	41	1001

address 1 of Wave 1, and the entry address number for each member of the household is the same as the current address number. Because they are assigned in Wave 1, the person numbers are in the 100 series and are numbered sequentially, beginning with 101.

- During Wave 2, the son joins the Army, moves into military barracks, and therefore leaves the SIPP sample.[29] The son's record, person number 104, will contain information (either

[29] Members of the armed forces are included in the SIPP sample only if they are living state-side in private housing. Those living overseas or in military barracks are not included in the SIPP sample universe.

imputed or provided by proxy) on his characteristics for the time in Wave 2 that he was still in the sample. If he does not return to the sample during the remainder of the panel, there will be no records for him beyond Wave 2.

- During Wave 3, the daughter marries and her husband moves into the household. The current address number where the mother, father, cousin, daughter, and son-in-law live remains the same because it is the same address. The son-in-law's entry address number is 11 because he first enters the SIPP sample at an address coded 11. The person number for the son-in-law is in the 300 series (301) because he joins the SIPP sample in Wave 3.
- During Wave 4, the daughter and son-in-law move into a new house. Their current address number changes to 41 to indicate that a new address has been established in Wave 4. Meanwhile, the cousin, who is over age 15, moves in with an uncle.[30] The cousin's current address number changes to 42 (i.e., the second household added into the SIPP sample in the fourth wave). The assignment of address number 41 to the daughter and 42 to the cousin is random. It could be the other way around. The uncle enters the SIPP sample and receives an address number of 42 and an entry address number of 42. The uncle's person number is in the 400 series (401) since he joins the survey in Wave 4.
- No changes in household composition are observed during Waves 5–9.
- During Wave 10, the daughter and son-in-law have a baby. This new sample member is assigned the sample unit ID of the daughter and son-in-law. The newborn's entry address is 41, since that is the current address ID of the daughter and son-in-law at the time of birth. The newborn's person number is 1001, reflecting the fact that the newborn came into the SIPP sample in Wave 10. Meanwhile, the cousin moves to Europe and therefore leaves the SIPP sample. The uncle, even though he did not move to Europe with the cousin, also leaves the SIPP sample because he no longer resides with an original SIPP sample member. Their records are no longer listed.

Table 12-18 displays this example again, but this table depicts how the HH-ADDID$_i$ variable changes over time to reflect the household composition changes. The table also illustrates the structure of the full panel data files.

There are two extremely rare occasions in which the original PP-ID, PP-ENTRY, and PP-PNUM values are modified:

1. The first occasion is when two separate sampling units, each containing original sample members, are merged, perhaps because of a marriage. In this situation, one of the original set of PP-ID and PP-ENTRY values is retained and the other set is changed to agree with the retained set. The person number values (PP-PNUM) of the changed set are modified further to be between 180 and 199, inclusive.

[30] In the 1993 Panel, all original sample members were followed, no matter what their ages. In all other panels, only people 15 years of age or older were followed when they moved to new addresses.

Table 12-18. Household Changes and Their Effects on the Household ID (HH-ADDIDi) Variable in the Longitudinal Research File

				HH-ADDID$_i$																			
				Wave 1				Wave 2				Wave 3				Wave 4				Wave 5			
	PP-	PP-		Month				Month				Month				Month				Month			
PP-ID	ENTRY	PNUM	Notes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
101111103	11	101	Father	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
101111103	11	102	Mother	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
101111103	11	103	Daughter	11	11	11	11	11	11	11	11	11	11	11	11	41	41	41	41	41	41	41	41
101111103	11	104	Son	11	11	11	11	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
101111103	11	105	Cousin	11	11	11	11	11	11	11	11	11	11	11	11	11	42	42	42	42	42	42	42
101111103	11	301	Son/law	0	0	0	0	0	0	0	0	0	11	11	11	41	41	41	41	41	41	41	41
101111103	42	401	Uncle	0	0	0	0	0	0	0	0	0	0	0	0	42	42	42	42	42	42	42	42
101111103	41	1001	Newborn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

				HH-ADDID$_i$																			
				Wave 6				Wave 7				Wave 8				Wave 9				Wave 10			
	PP-	PP-		Month				Month				Month				Month				Month			
PP-ID	ENTRY	PNUM	Notes	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
101111103	11	101	Father	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
101111103	11	102	Mother	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
101111103	11	103	Daughter	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41
101111103	11	104	Son	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
101111103	11	105	Cousin	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	0	0	0	0
101111103	11	301	Son/law	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41
101111103	42	401	Uncle	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	0	0	0	0	0
101111103	41	1001	Newborn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	41	41	41	41

2. The second occasion is when a household splits into two new households (in which each new household gains a new sample person) and later the households recombine. For example, assume that a married couple separate in Wave 3, each moving in with a sibling. Both siblings are assigned a person number of 301, because they entered the sample in Wave 3 at different addresses (thus, HH-ADDID$_i$ = 31 and 32). If the husband and wife reunite in Wave 6, and bring the siblings with them, one sibling's person number would be changed. In this case, one of the siblings would have a person number of 301 and the other would have a person number of 680 (or some number between 680 and 699, inclusive).

Because a record in the longitudinal research file describes the person throughout the entire panel and because the sample unit ID (PP-ID) cannot change on this record, each person in a merged household whose ID values were changed is assigned two full panel records. The first record contains the original ID information of the person before the merge and identifies the person as having exited the sample at the time of the merge. The second record contains the new ID information and identifies the person as having entered the sample at the time of the merge. There is no way to link the two records in the longitudinal research files.[31]

Identifying Program Units

Besides household and family composition data, the longitudinal research files contain detailed information about participation in health insurance and various government transfer programs. For most programs, three characteristics are recorded (Table 12-19):

1. Whether the person is covered;
2. Who received the income or benefit; and
3. The amount of the income or benefit.

The coverage variables identify whether the income or benefit covers that person in month i. In other words, when a person is flagged as covered by food stamps (FOODSTMP$_i$ = 1), the person either received the benefits directly (because he or she was the authorized food stamp recipient) or indirectly (because he or she was in the same program unit as the authorized recipient). The coverage variables also allow users to determine each person's membership in each program unit. That is useful because program units often exclude some members of the family or household.[32] Also, as with households and families, membership in program units can change from one month to the next. For that reason, program unit membership and characteristics of the unit should be evaluated for each month.

[31] If needed, this information can be merged from the core wave files. Chapters 10 and 13 provide details.

[32] In the 1984 and 1985 Panels, coverage for the Women, Infants, and Children (WIC) nutrition program was imputed to children under 6 years old if their mother reported participation in the WIC program. Beginning with the 1986 Panel, WIC coverage has been assessed directly for all sample members.

Table 12-19. Variables Describing Participation in Government Transfer Programs and Health Insurance Programs in the 1990–1993 Longitudinal Research Files

Program	Coverage	Authorized Recipient	G1 Source Code	Amount
Social Security	SOC-SEC	SS-PIDX	1	Locate one of the amount variables: G1AMT1–G1AMT10, using the corresponding source variables: G1SRC1–G1SRC10
Railroad Retirement	RAILROAD	RR-PIDX	2	
Federal Supplemental Security Income	—	—	3	
Veteran's Benefits	VETS	VA-PIDX	8	
Aid to Families with Dependent Children	AFDC	AFDCPIDX	20	
General Assistance	GEN-ASST	GA-PIDX	21	
Foster Child Care	FOST-KID	FOSTPIDX	23	
Other Welfare	OTH-WELF	OTH-PIDX	24	
WIC Benefits	WICCOV	WIC-PIDX	25	
Food Stamps	FOODSTMP	FS-PIDX	27	
Medicare	CARECOV	—	—	
Medicaid	CAIDCOV	—	—	
CHAMPUS	CHAMP	—	—	

The authorized recipient variables identify the people who actually received the income or benefit for the people in their program units. In the longitudinal research files, those variables do not use the entry address and person number values. Instead, they use the sequence number of the person within the sample unit (PP-RCSEQ) to identify authorized recipients. In other words, the authorized food stamp recipient is the person for whom FS-PIDX$_i$ in month i equals PP-RCSEQ.

Individuals who are members of a common program unit in a given month (i) can be identified by using the sample unit ID (PP-ID), the person's interview status in month i (PP-MIS$_i$), and the authorized recipient variable in month i. For example, members of a common food stamp unit in month i are those with PP-MIS$_i$ of 1 and common values of PP-ID (a value that does *not* change from month to month) and FS-PIDX$_i$ (a value that *does* change from one month to the next). The SIPP longitudinal research files do not include authorized recipient variables for Medicare and SSI programs.[33]

There are some exceptions to the rules:

- Social Security, Railroad Retirement, WIC, and AFDC can offer benefits solely to children. When that happens, an adult will receive the income on behalf of the children. The adult, therefore, is flagged as the authorized recipient and the income amounts appear on the record of the adult. The adult authorized recipient, however, is not flagged as being covered by the program. The children are flagged as covered.

[33] In effect, each person covered by these two programs is an authorized recipient, and the program units are the people themselves.

- Most SSI recipients are elderly and disabled adults, but they can also be children with disabilities.[34] Even so, the SSI amount is recorded on an adult's record, not on the child's record. Unlike the core wave files, the longitudinal research files have no coverage variable indicating whether or not the child, adult, or both, were covered. If needed, this information can be merged from the core wave files. Chapter 13 provides a detailed discussion of merging SIPP files.
- The medical insurance variables simply reflect who is enrolled in which type of program. There are no associated amount variables.

These rules and exceptions are illustrated in Table 12-20. The household contains one AFDC unit and two food stamp units. The mother is covered by Social Security and SSI. The mother of the (disabled) child receives SSI on behalf of her child. The grandchild receives WIC. Everyone in the household is enrolled in Medicaid. The coverage variables are set to 2 whenever the person is *not* covered by the particular program. The indicators for the authorized recipients do *not* use the PP-ENTRY and PP-PNUM values. Instead, they are based on the "line number" of the authorized recipient on the household roster. That is very different from the indicators used on the core wave files.

Using the Unearned Income Variables

To save space, the Census Bureau organizes the unearned income variables differently in the longitudinal research files than in the core wave files. As shown in Table 12-21, 10 variables on each person's record identify up to 10 different sources of unearned income (G1SRC1–G1SRC10). For each source identified, there is a corresponding amount variable ($G1AMT1_i$–$G1AMT10_i$). Income amounts are recorded with monthly resolution. The person in Table 12-21 periodically receives $500 in federal SSI and $125 in food stamps. The person does not receive any other source of unearned income.

When using these fields, analysts often find it helpful to realign the unearned income into new income-specific variables.[35]

[34] In the 1990s, the definition of qualifying disabling conditions was expanded. That change in definition resulted in a rapid expansion of the child SSI caseload.

[35] For example, Table 12-22 includes monthly variables for SSI and food stamps that were created by using the algorithm in Figure 12-5.

Table 12-20. Example of Program Units, Coverage, and Benefit Amounts in the Longitudinal Research Files

Variable	Mother	Daughter #1	Daughter #1's Son	Daughter #2	Spouse of Daughter #2
PP-PNUM	101	102	103	104	105
PP-RCSEQ	1	2	3	4	5
AGE_i	70	21	4	25	26
AFDC					
$AFDC_i$	2	1	1	2	2
$AFDCPIDX_i$	0	2	2	0	0
Food Stamps					
$FOODSTMP_i$	2	1	1	1	1
$FS\text{-}PIDX_i$	0	2	2	4	4
SSI					
This only appears in the General Amounts (G1) section.					
WIC					
$WICCOV_i$	2	2	1	2	2
$WIC\text{-}PIDX_i$	0	2	2	0	0
Medicaid					
$CAIDCOV_i$	1	1	1	1	1
Social Security					
$SOC\text{-}SEC_i$	1	2	2	2	2
General (G1) Sources and Amounts					
G1SRC1	3	20	0	27	0
$G1AMT1_i$ ($)	188	123	0	130	0
G1SRC2	1	27	0	0	0
$G1AMT2_i$ ($)	470	160	0	0	0
G1SRC3	0	3	0	0	0
$G1AMT3_i$ ($)	0	122	0	0	0
G1SRC4	0	25	0	0	0
$G1AMT4_i$ ($)	0	30.12	0	0	0

[a] These codes are explained in the next section of text.

Income Topcoding

The Census Bureau topcodes each income variable to protect against the possibility that a user might identify a SIPP respondent with very high income.[36] While the data dictionary indicates a topcode of $33,332 for monthly income, that is also the income topcode for the wave. That topcode is, therefore, rarely used for a month. In most cases, the monthly income is topcoded at $8,333, which actually represents $8,333 or more. Individual amounts above $8,333 may occasionally be shown if the respondent's income varied considerably from month to month

[36] New topcoding procedures are being implemented with the 1996 Panel. When a longitudinal research file for the 1996 Panel is available, this discussion will be revised to describe those new procedures. At present, users should note that this description does *not* pertain to the core wave files from the 1996 Panel.

Table 12-21. Unearned Income in the Longitudinal Research Files

		PP-MIS																			
		Wave 1				Wave 2				Wave 3				Wave 4				Wave 5			
		Month				Month				Month				Month				Month			
Variable		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
PP-ID	7887																				
PP-PNUM	102																				
PP-MIS		1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	0	0	0	0
G1SRC1	3																				
G1AMT1 ($)		500	500	500	500	0	0	0	500	500	500	500	500	0	0	0	0	0	0	0	0
G1SRC2	27																				
G1AMT2 ($)		0	0	0	0	0	0	0	125	125	125	125	0	0	0	0	0	0	0	0	0
G1SRC3	0																				
G1AMT3 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC4	0																				
G1AMT4 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC5	0																				
G1AMT5 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC6	0																				
G1AMT6 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC7	0																				
G1AMT7 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC8	0																				
G1AMT8 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC9	0																				
G1AMT9 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC10	0																				
G1AMT10 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 12-21. Unearned Income in the Longitudinal Research Files (*continued*)

		PP-MIS																		
		Wave 6			Wave 7			Wave 8				Wave 9			Wave 10					
		Month			Month			Month				Month			Month					
Variable		**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**29**	**40**
PP-ID PP-PNUM PP-MIS	7887 102	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC1 G1AMT1($)	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC2 G1AMT2($)	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC3 G1AMT3($)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC4 G1AMT4($)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC5 G1AMT5($)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC6 G1AMT6($)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC7 G1AMT7($)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC8 G1AMT8($)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC9 G1AMT9($)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC10 G1AMT10 ($)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 12-22. User-Created SSI and FSP Variables Using the Unearned Income Variables in the Longitudinal Research Files

		PP-MIS																			
		Wave 1				Wave 2				Wave 3				Wave 4				Wave 5			
		Month				Month				Month				Month				Month			
Variable		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
PP-ID	7887																				
PP-PNUM	102																				
PP-MIS		1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	0	0	0	0
G1SRC1	3																				
G1AMT1 ($)		500	500	500	500	0	0	0	500	500	500	500	500	0	0	0	0	0	0	0	0
G1SRC2	27																				
G1AMT2 ($)		0	0	0	0	0	0	0	125	125	125	125	0	0	0	0	0	0	0	0	0
G1SRC3	0																				
G1AMT3 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC4	0																				
G1AMT4 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC5	0																				
G1AMT5 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC6	0																				
G1AMT6 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC7	0																				
G1AMT7 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC8	0																				
G1AMT8 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC9	0																				
G1AMT9 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC10	0																				
G1AMT10 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SSI ($)		500	500	500	500	0 [a]	0	0	500	500	500	500	500	–99	–99	–99	–99	–99	–99	–99	–99
FSP ($)		0	0	0	0	0	0	0	125	125	125	125	0	–99	–99	–99	–99	–99	–99	–99	–99

[a] In SAS, the unassigned values would have a "system missing" value displayed as a ".".

Table 12-22. User-Created SSI and FSP Variables Using the Unearned Income Variables in the Longitudinal Research File (*continued*)

		PP-MIS																			
		Wave 6				Wave 7				Wave 8				Wave 9				Wave 10			
		Month				Month				Month				Month				Month			
Variable		21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
PP-ID	7887																				
PP-PNUM	102																				
PP-MIS		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC1	3																				
G1AMT1 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC2	27																				
G1AMT2 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC3	0																				
G1AMT3 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC4	0																				
G1AMT4 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC5	0																				
G1AMT5 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC6	0																				
G1AMT6 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC7	0																				
G1AMT7 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC8	0																				
G1AMT8 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC9	0																				
G1AMT9 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1AMT10 ($)																					
SSI ($)		–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99
FSP ($)		–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99

Figure 12-5. Creating Monthly Food Stamp and SSI Income Variables from the Unearned Income Variables in the Longitudinal Research Files

```
For each person:
      /*
            This step is not needed in SAS
      */
      For each month (index = mo):
            If PP-MIS (mo) = 1 Then do
                  SSI(mo) = 0
                  FSP(mo) = 0
            End If PP-MIS (mo) = 1
            Else do
                  SSI(mo) = -99
                  FSP(mo) = -99
            End Else
      End month loop
      /*
            Begin here for SAS
      */
      For each G1SRC (index=i):
            If G1SRC(i)=3 Then do
                  For each month (index=mo)
                        If PP-MIS (mo) = 1 Then do SSI(mo)=G1AMT(i,mo)
                        End If PP-MIS (mo) = 1
                  End month loop
            End If G1SRC(i)=3
            Else if G1SRC(i)=27 Then do
                  For each month (index=mo)
                        If PP-MIS (mo) = 1 Then do FSP(mo)=G1AMT(i,mo)
                        End If PP-MIS (mo) = 1
                  End month loop
            End if G1SRC(i)=27
      End G1SRC loop
```

within a wave. For example, if a respondent's income from a single job was concentrated in only one of the four reference months, a figure as high as $33,332 could be shown.

Summary income variables on the person, family, and household records are simply the sums of the component variables after they have been topcoded. The summary variables are not independently topcoded. Thus, a person with high income from several sources (multiple jobs, businesses, property) could have aggregate monthly income well over the topcode for each source, and yet the data could still be greatly understating the person's true income.

As shown in Table 12-23, person 101 has wages topcoded. The person received considerably more money in December than in the other months. Also, total family income and total household income are the sum of the income amounts (in this case, WS-ERN-AMT1$_i$ + G1AMT1$_i$) after they have been topcoded.

Table 12-23. Example of Topcoding in the Longitudinal Research Files

Person Number (PP-PNUM)	Calendar Month	Household Total Income (HH-INC$_i$)	Family Total Income (FF-INC$_i$)	Wages (WS-ERN-AMT1$_i$)	Child Support Payments (G1AMT1$_i$)
101	10	$ 9,333	$ 9,333	$ 8,333	$1,000
101	11	$ 9,333	$ 9,333	$ 8,333	$1,000
101	12	$13,123	$13,123	$12,123[a]	$1,000
101	01	$ 5,793	$ 5,793	$ 4,543	$1,250

[a] This figure can exceed the nominal *monthly* topcode of $8,333 because the person's total earnings for the *wave* were below $33,332.

Using Allocation (Imputation) Flags

As described in Chapter 4, the Census Bureau often imputes information when a person does not respond to the survey or to a particular question. Two sources identify whether information has been imputed:

1. Beginning with the 1991 Panel, all data for a wave are imputed if a person was not successfully interviewed in one wave but had complete information (from either a successful interview or a proxy interview) in the two adjacent waves. In those cases, the value of WAVFLG will be greater than zero and INTVW will be 3 or 4.
2. A variable of interest may be imputed. In the longitudinal research files, allocation (imputation) flags are included for the earned income, asset income, and unearned (transfer) income variables.

Other variables are also subject to editing and imputation. The edit and imputation procedures used for the longitudinal research files differ from those used for the core wave files. The procedures used for the longitudinal research files make use of the full set of longitudinal data for a person. Because the core wave files are processed individually, the edit and imputation procedures applied to those files have, at most, 4 months of observations for a person. The procedures applied to the core wave files make greater use of cross-observation imputation methods than do those applied to the longitudinal research files.[37]

Using Weights

The full panel longitudinal research files include the calendar year weights (FNLWGTs) and the full panel weight (PNLWGT). The number of calendar year weights depends on the duration of

[37] The edit and imputation procedures applied to the core wave files from the 1996 Panel make greater use of retrospective information than procedures used in earlier panels. See Chapters 4 and 10 for details.

the panel; the number varies from one calendar year weight for the 1989 Panel to three calendar year weights for the 1993 Panel. When the 1996 full panel file is available, it will have four calendar year weights.

The source and accuracy statements that accompany all SIPP full panel files ordered from the Census Bureau provide suggestions on how to use the weight variables in those files. Also, Chapter 8 of this Guide contains a full discussion of how to use weights in full panel files.

Identifying States

The longitudinal research file contains a variable (GEO-STE) that identifies 41 individual states and the District of Columbia; the nine other states are suppressed into three groups:

1 Maine, Vermont;

2. Iowa, North Dakota, South Dakota; and

3. Alaska, Idaho, Montana, Wyoming.

Even though it is possible to identify most states, the SIPP sample was not designed to be representative at the state level and should not be used to produce direct state-level estimates. The state variable is included on the public use files to allow examination of how state-level characteristics affect national estimates. For example, a user could apply the state-specific eligibility criteria for a means-tested program in order to arrive at a national estimate of the number of people eligible for the program. Because some states are not uniquely identified, some method of allocating the state-specific eligibility rules to sample persons in those states would need to be devised.

Identifying Metropolitan Areas

The longitudinal research files do not contain any variables identifying metropolitan areas. Analysts who need this information should merge it from the core wave files. Chapter 11 provides details about how to use the variables identifying metropolitan areas. Chapter 13 provides instructions for merging data from multiple SIPP public use files.